阿拉伯数字笔迹特征提取分析及其鉴定应用初探

【摘要】本课题以阿拉伯数字笔迹作为研究对象，在前期收集样本的基础上，对 0 至 9 十个阿拉伯数字提取一定量的特征（包括上下左右比例以及不同部位的曲率、夹角、倾斜角），建立阿拉伯数字笔迹特征数据库，将每个阿拉伯数字量化成一定的数据；计算了同一人之间、不同人之间书写的阿拉伯数字笔迹样本的吻合度、不同组合数目的阿拉伯数字笔迹组合的样本吻合度及不同种类特征间的相对离散性；通过有效性分析和测试，说明了研究方法的有效性、可重复性；最后综合分析，对阿拉伯数字笔迹进行系统的梳理及其在司法鉴定领域应用的探讨。课题结果表明，同一个人书写的样本和不同人书写的样本的吻合度存在显著差异，可以很好地区分开来；不同形式的特征的稳定性不同，角度的稳定性最好，区分度最大，曲率、上下左右比例也有较好的稳定性，次之于角度。通过更深入的研究，本课题使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法可为日后真正的鉴定提供参考和理论支持。

关键词：阿拉伯数字、笔迹、特征提取、司法鉴定

Abstract

The project focuses on the Arabic numbers’ handwritings. On the basis of pre-collected samples, the author extracted a certain amount of features (including curvature, angle, tilt angle, proportion) and established the feature database of the Arabic numbers’ handwritings. The author calculated the accordance rate of Arabic numbers written by the same people, the accordance rate of Arabic numbers written by different people, the accordance rate of Arabic numbers in different combinations, the relative discreteness of different features, conducted effectiveness analysis to verify the effectiveness and repeatability of the research methods used. Through testing, the author determined the stability of the selected features and the effectiveness of the research method. Finally, the author did comprehensive analysis of Arabic numbers’ handwritings and discussed its application in the field of forensic science. The result shows that the accordance rate of the same people and of different people has significant difference and can be distinguished. Different features have different stability and the stability of them vary. The angle has the greatest stability while the curvature and proportion follow it. Through further research, the research method used can be applied in real forensic science field in future.

Key words: Arabic numbers, handwriting, features extraction, forensic science

目录

1 引言 3

1.1 课题由来 3

1.2 本领域相关研究和司法鉴定实践状况 3

2 研究目的与方法 4

2.1 研究目的 4

2.2 研究方法 4

3 阿拉伯数字笔迹特征提取分析及其鉴定应用 5

3.1 阿拉伯数字笔迹特征库 5

3.2 阿拉伯数字笔迹特征吻合度分析 7

3.3.1 同一人书写样本吻合度比较分析 8

3.3.2 不同人书写样本吻合度比较分析 9

3.3.3 阿拉伯数字笔迹特征吻合度综合比较分析 10

3.3.4 阿拉伯数字笔迹组合吻合度分析 11

3.4 阿拉伯数字笔迹不同特征相对离散性比较 13

3.5 本鉴定方法有效性分析 16

3.5.1 有效性分析：增加样本比较特征数 16

3.5.2 有效性分析：增加提取样本特征并进行比较的次数 20

3.6 测试 21

3 结论与讨论 22

3.1 结论 22

3.2 课题讨论 22

3.3 创新点 23

3.4 课题展望 23

致谢 24

参考文献 24

附：收获与体会 25

1 引言

当今社会不断加速发展，阿拉伯数字作为整个文字系统中重要的一部分，成为了各类经济纠纷案件中的一个焦点，常常需要对其书写的真伪性展开追究判断。在各类案件中，如借条收据类和落款日期数字类中，阿拉伯数字都起到了至关重要的作用。目前我国在相关鉴定中积攒了一定的经验，但在系统的研究和理论方面还是有所欠缺的。总体来说，关于阿拉伯数字方面的鉴定主要还是基于主观经验的积累，并没有强有力的科学理论作为支撑。阿拉伯数字的书写和汉字一样，有一定的动力定型，人们书写的动作习惯在一定时间内保持相对稳定，则是可以对阿拉伯数字进行鉴定的前提。

针对现今阿拉伯数字笔迹鉴定的主观经验性，本课题以阿拉伯数字笔迹作为研究对象，对 0 至 9 十个阿拉伯数字笔迹进行特征提取并展开统计分析研究。在收集到一定数量的样本后，对每个阿拉伯数字提取一定量的特征，简单而言就是数字有弯曲的部分（例如 0、2、3、5、6、8、9）提取曲率作为特征、线条较笔直光滑的转折处（例如 2、4、5、7）提取倾斜的角度为特征、也包括上下左右比例等，建立阿拉伯数字笔迹特征数据库，把每个阿拉伯数字量化成多个数据；计算同一人之间、不同人之间书写的阿拉伯数字笔迹样本的吻合度、不同组合数目的阿拉伯数字笔迹组合的样本吻合度、不同种类特征间的相对离散性；通过本鉴定方法的有效性分析和测试，说明了研究方法的有效性、可重复性。最后综合分析，得出更科学准确的结论，对阿拉伯数字笔迹进行系统的梳理和科学的总结，以此能为今后阿拉伯数字笔迹鉴定实践过程中提供一定的理论依据支撑和可供参考的新思路，这一点是本课题最大的实际应用。除此之外，这项研究项目也能让大家对这个平时不是很熟知的领域有一些新的了解。

1.1 课题由来

在观看各类案件的电视节目过程中，常常会了解到各类经济纠纷案件，其中的焦点往往集中于借条收据类和落款日期数字类中阿拉伯数字书写的真伪性，我也对到底是如何进行系统的阿拉伯数字笔迹鉴定产生了兴趣。但经过查阅文献和走访有关部门的笔迹鉴定专家却发现，关于阿拉伯数字笔迹的判断往往是基于鉴定人的主观经验，系统的定量研究是有所欠缺的，并没有强有力的科学理论作为支撑。但阿拉伯数字的书写和汉字一样，有一定的动力定型，书写的动作习惯在一定时间内保持相对稳定，有统一认定的前提。这一点引起了我的思考，想尝试对阿拉伯数字笔迹从数学上面进行特征提取与定量的分析，建立系统的分析方法。

1.2 本领域相关研究和司法鉴定实践状况

目前，系统地研究阿拉伯数字笔迹的文献并不多见，主要为概述型，以理论探讨居多，列举了一些阿拉伯数字笔迹形态方面的特征，如起笔的顺逆时针、连笔、形状等，但都只是一些定性的判断方法。其中较深入的一篇是西南政法大学的一篇硕士论文《阿拉伯数字笔迹鉴定研究》，但该文对阿拉伯数字笔迹特征的描述只是停留在了形态方面，并没有用数学的特征定量地来描述这个问题。

在采访司法鉴定领域工作的鉴定人员时，他们坦诚地告知现在关于阿拉伯数字笔迹鉴定方面的工作很大部分依赖于经验判断，没有采取定量的判断方法，非常需要理论方面的支持。

由此知道，现今针对阿拉伯数字笔迹的鉴定，是基于形态分析的较为主观经验性的东西，仍需要进一步的补充和探讨，阿拉伯数字笔迹鉴定方面的研究还有很大的发展空间。

2 研究目的与方法

2.1 研究目的

研究主要针对现今阿拉伯数字笔迹鉴定的主观经验性，以阿拉伯数字笔迹作为研究对象，对 0 至 9 十个阿拉伯数字笔迹的特征进行提取并展开统计分析研究。认为能用数学的方法将不同人书写的阿拉伯数字区分开来，通过有效性分析和测试证明研究方法的有效性、可重复性。最后综合分析，对阿拉伯数字笔迹进行系统的梳理和科学的总结，并进行司法鉴定领域的探讨，以此能为今后阿拉伯数字笔迹鉴定实践过程中提供一定的理论依据支撑和可供参考的新思路。

2.2 研究方法

1 前期样本收集

前期收集 100 份不同人书写的 0 至 9 十个阿拉伯数字的样本，从中观察得出不同人差异较大、可供提取的阿拉伯数字笔迹特征，建立阿拉伯数字笔迹特征库。

2 第二轮样本收集

收集一定数量的不同人书写的三遍 0 至 9 十个阿拉伯数字的样本，在书写的样纸上大致画出长宽范围限制，防止书写太过随意大小经处理后放大或缩小图片清晰度受影响。收集后扫描入电脑用 photoshopCS3 进行除噪、像素清晰化、裁剪至统一尺寸大小等预处理。

3 特征提取

用 ImageJ 测量软件对所有收集的阿拉伯数字笔迹样本进行特征提取并记录。

4 统计分析

对所得所有数据进行整合，展开统计分析，计算同一个人所书写样本和不同人所书写样本之间的吻合度、不同数量的阿拉伯数字笔迹组合的吻合度并进行比较分析。

用变异系数来计算判断阿拉伯数字笔迹不同种类特征的相对离散性，分析阿拉伯数字笔迹不同种类特征的稳定性。

5 有效性分析

有效性分析分为两方面：

1 对同一个人书写的每个阿拉伯数字，不断增加进行比较的特征的数目，观察吻合度是否趋近于并稳定在一定范围，并得出使每个阿拉伯数字的吻合度开始趋于稳定的特征数。

2 选取特征较多、具有代表性的阿拉伯数字，不断增加提取样本特征并进行比较的次数，观察吻合度是否趋近于并稳定在一定范围内。

从而证明本课题中使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法具有有效性、可重复性，在日后真正的鉴定领域内可以应用。

6 测试

取其中一个阿拉伯数字作为比较的对象，任选两个人，令其分别书写该数字三遍，提取并计算出关于该数字特征的平均值作为其关于该数字的特征的数据，计算各个特征的标准差，进一步验证选取的特征是否稳定具有科学性以便进行调整。再令两个人分别又书写六遍该数字，用比较吻合度的方法判断这些数字分别是谁书写的，观察得出的正确率，以此验证本课题中使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法具有有效性，在日后真正的鉴定领域内可以应用。

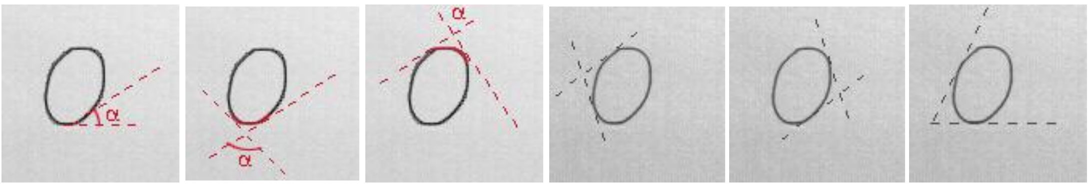
3 阿拉伯数字笔迹特征提取分析及其鉴定应用

3.1 阿拉伯数字笔迹特征库

根据前期收集 100 份不同人书写的 0 至 9 十个阿拉伯数字的样本，从中观察得出不同人差异较大、可供提取的阿拉伯数字笔迹特征，建立阿拉伯数字笔迹特征库。其中，简单而言就是数字有弯曲的部分（例如 0、2、3、5、6、8、9）提取曲率作为特征、线条较笔直光滑的转折处（例如 2、4、5、7）提取倾斜的角度为特征、也包括各部分的上下左右比例等。

阿拉伯数字“0”

阿拉伯数字“0”整体较平滑，提取与水平面的倾斜角度作为特征，提取多处曲率作为特征。



阿拉伯数字“1”

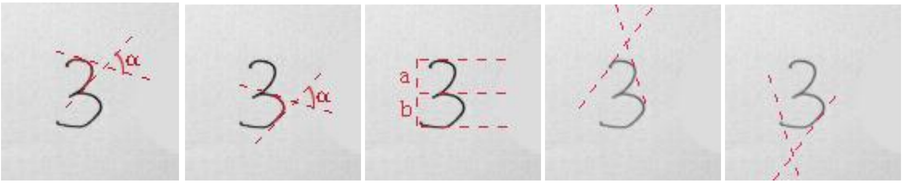
阿拉伯数字“1”整体特征较单调，主要是与水平面和竖直面的夹角。

阿拉伯数字“2”

阿拉伯数字“2”提取上方、右方弯曲处的曲率、与水平面的倾斜角、起笔与转折处的夹角作为特征。

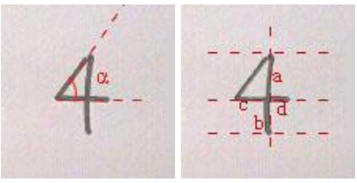
阿拉伯数字“3”

阿拉伯数字“3”提取多处曲率、上下比例作为特征。



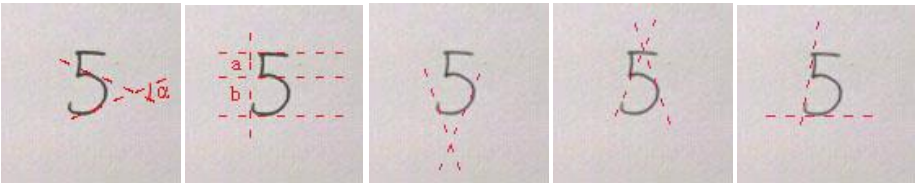
阿拉伯数字“4”

阿拉伯数字“4”提取第一处转折处的夹角、上下左右比例作为特征。



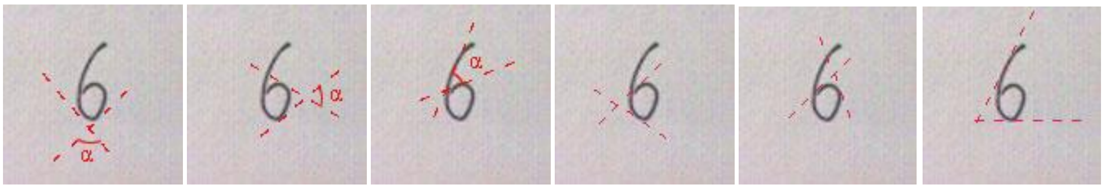
阿拉伯数字“5”

阿拉伯数字“5”提取多处曲率、上下比例、起笔与收笔处连线和水平面的夹角作为特征。



阿拉伯数字“6”

阿拉伯数字“6”提取多处曲率，上半部分延长线与水平面的倾斜角作为特征。

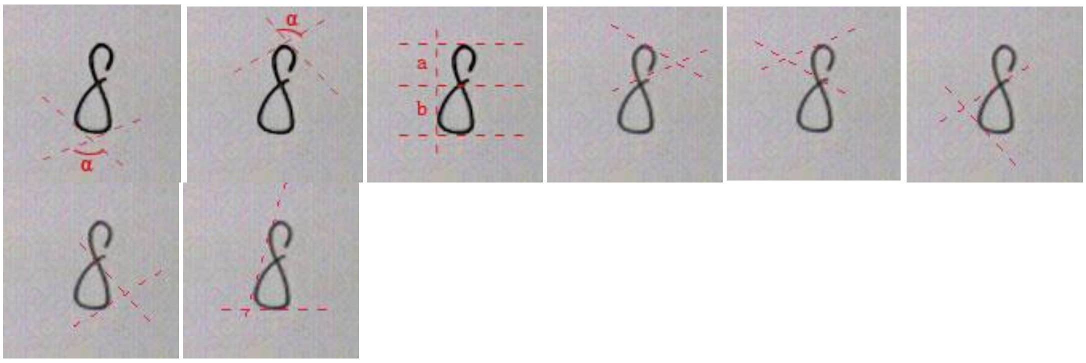


阿拉伯数字“7”

阿拉伯数字“7”提取与水平面倾斜角、第一笔与第二笔转折处的夹角、起笔与收笔处的夹角作为特征。

阿拉伯数字“8”

阿拉伯数字“8”提取多处曲率、上下比例、左上部分延长线与水平面倾斜角作为特征。



阿拉伯数字“9”

阿拉伯数字“9”提取上方，左方，右方曲率、与水平面倾斜角作为特征。

3.2 阿拉伯数字笔迹特征吻合度分析

收集的样本中，每个人书写了三遍 0 至 9 十个阿拉伯数字，从中随机选取 2 组样本进行探究。

关于特征吻合的条件定义为：相对误差 α < 0.2

$$\text{相对误差}\ \alpha = \left|\frac{x_i - x_j}{X_h}\right| \qquad X_h = \frac{X_i + X_j}{2}$$

其中，X_i、X_j 分别为正在进行一一比较的特征的数据，X_h即X_i、X_j的算术平均值。

关于特征吻合度的定义为：

$$吻合度=\frac{吻合的特征数}{比较的总特征数}\times 100\%$$

3.3.1 同一人书写样本吻合度比较分析

在选取出的两大组样本中，将每大组其中的三小组 0 至 9 十个阿拉伯数字的样本分别记为 1.1、1.2、1.3 和 2.1、2.2、2.3，计算同一人书写样本吻合度的时候，组内两两进行比较，即 1.1、1.2、1.3 三组样本两两比较计算吻合度，2.1、2.2、2.3 三组样本两两比较计算同一人书写样本的吻合度。

表 1、表 2 中，横行代表进行两两比较的组，纵列代表进行比较的阿拉伯数字，列出的数据表示这种情况下的吻合度（%）。

表 1：阿拉伯数字笔迹样本 1 特征吻合度

	1.1,1.2	1.1,1.3	1.2,1.3
0	83.33	66.67	66.67
1	100	100	100
2	100	80	80
3	80	80	60
4	100	100	100
5	60	100	80
6	66.67	83.33	83.33
7	100	100	100
8	75	87.5	75
9	80	80	80

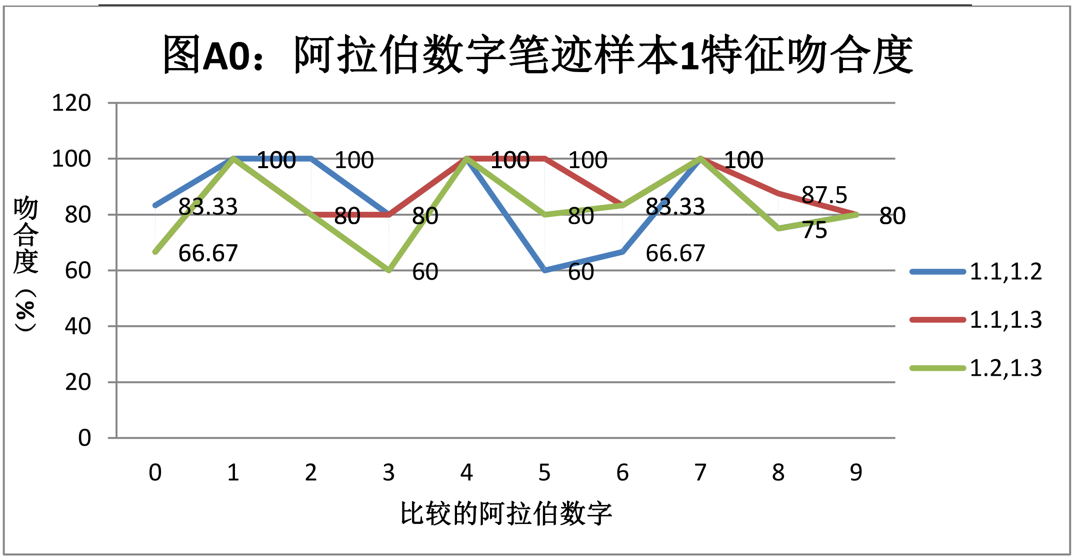
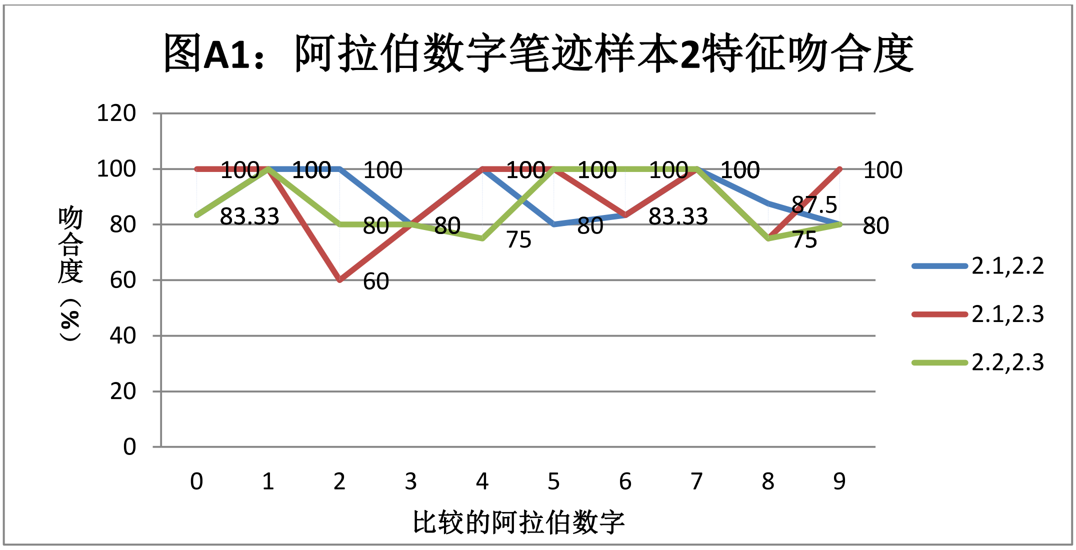


表 2：阿拉伯数字笔迹样本 2 特征吻合度

	2.1，2.2	2.1，2.3	2.2,2.3
0	83.33	100	83.33
1	100	100	100
2	100	60	80
3	80	80	80
4	100	100	75
5	80	100	100
6	83.33	83.33	100
7	100	100	100
8	87.5	75	75
9	80	100	80



由表 1、2 及图 A0、A1 可知，同一个人书写的样本中，样本特征具有一定的稳定性，样本特征的吻合度大致稳定在 80%—100%之间。

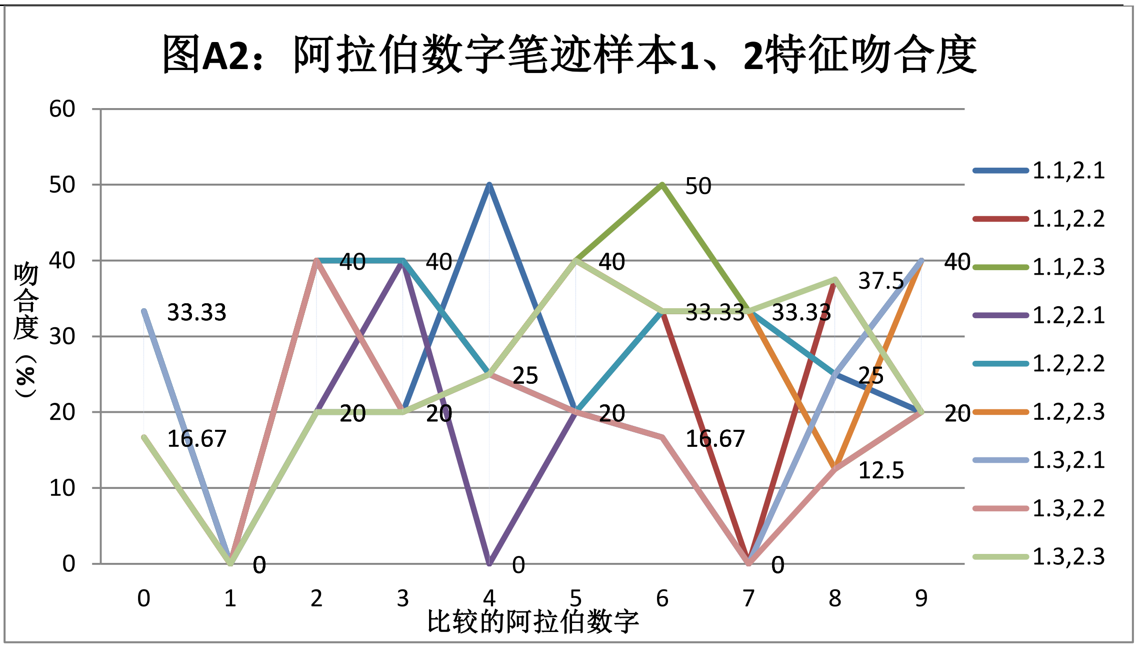
3.3.2 不同人书写样本吻合度比较分析

在选取出的两大组样本中，将每大组其中的三小组 0 至 9 十个阿拉伯数字的样本分别记为 1.1、1.2、1.3 和 2.1、2.2、2.3，计算不同人书写样本吻合度的时候，两大组的三小组数据分别进行两两比较，即 1.1、1.2、1.3 三组样本分别与 2.1、2.2、2.3 三组样本两两比较计算不同人书写样本的吻合度。

表 3 中，横行代表进行两两比较的组，纵列代表进行比较的阿拉伯数字，列出的数据表示这种情况下的吻合度（%）。

表 3：阿拉伯数字笔迹样本 1、2 特征吻合度

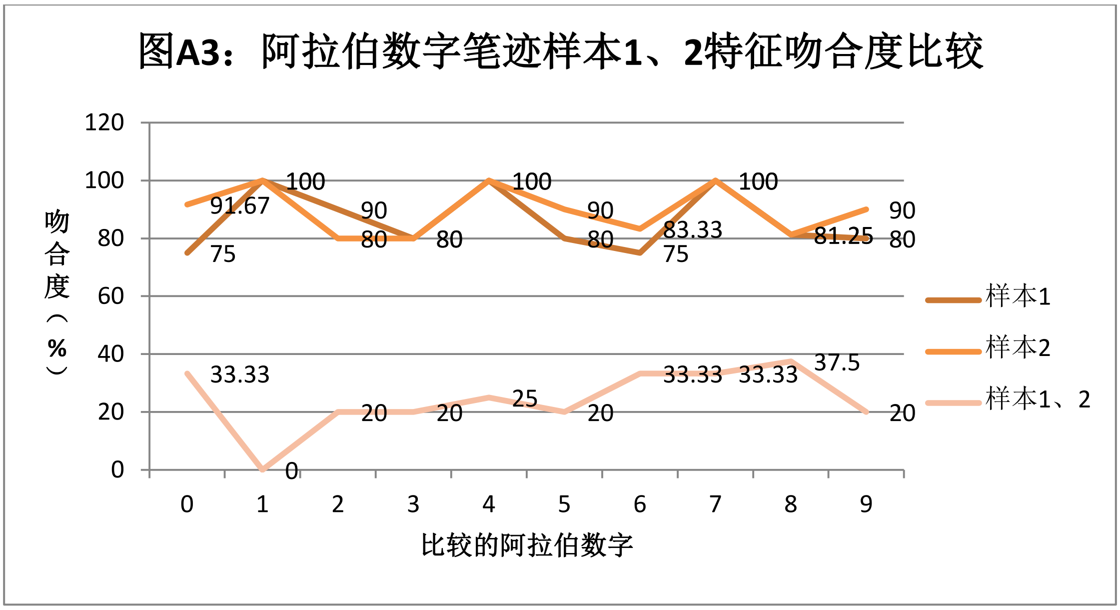
	1.1,2.1	1.1,2.2	1.1,2.3	1.2,2.1	1.2,2.2	1.2,2.3	1.3,2.1	1.3,2.2	1.3,2.3
0	33.33	16.67	33.33	33.33	33.33	16.67	33.33	16.67	16.67
1	0	0	0	0	0	0	0	0	0
2	40	40	20	20	40	40	20	40	20
3	20	40	20	40	40	20	20	20	20
4	50	25	25	25	0	25	25	25	25
5	20	40	40	20	20	40	20	20	40
6	33.33	33.33	50	16.67	33.33	33.33	16.67	16.67	33.33
7	33.33	0	33.33	0	33.33	33.33	0	0	33.33
8	25	37.5	12.5	25	25	12.5	25	12.5	37.5
9	20	20	20	40	40	40	40	20	20



由表 3 及图 A2 可知，不同人书写的样本中，样本特征的吻合度大致稳定在 20%—40%之间，较同一人书写的样本特征的吻合度来看，吻合度明显降低。

3.3.3 阿拉伯数字笔迹特征吻合度综合比较分析

将样本 1 的 3 组同一人书写样本吻合度、样本 2 的 3 组同一人书写样本吻合度、样本 1，2 的 9 组不同人书写样本的吻合度求算术平均值，分别得到样本 1 自己的书写样本吻合度、样本 2 自己的书写样本吻合度、样本 1,2 之间的书写样本吻合度，得到下图。



由图 A3 可知，同一个人书写的样本中，吻合度较高，吻合度大致稳定在 80%—100%之间，不同人书写的样本中，吻合度较低，吻合度大致稳定在 20%—40%之间，二者间有显著差异，可以区分开来。

3.3.4 阿拉伯数字笔迹组合吻合度分析

对同一个人书写的阿拉伯数字，将不同的阿拉伯数字组合在一起，不断增加组合在一起的阿拉伯数字的组合数目，由两个两个组合增至九个九个组合（即组合数目由 2 增至 9），观察吻合度的变化。（如组合数目为 4 指同时比对四个阿拉伯数字，观察此时的吻合度）

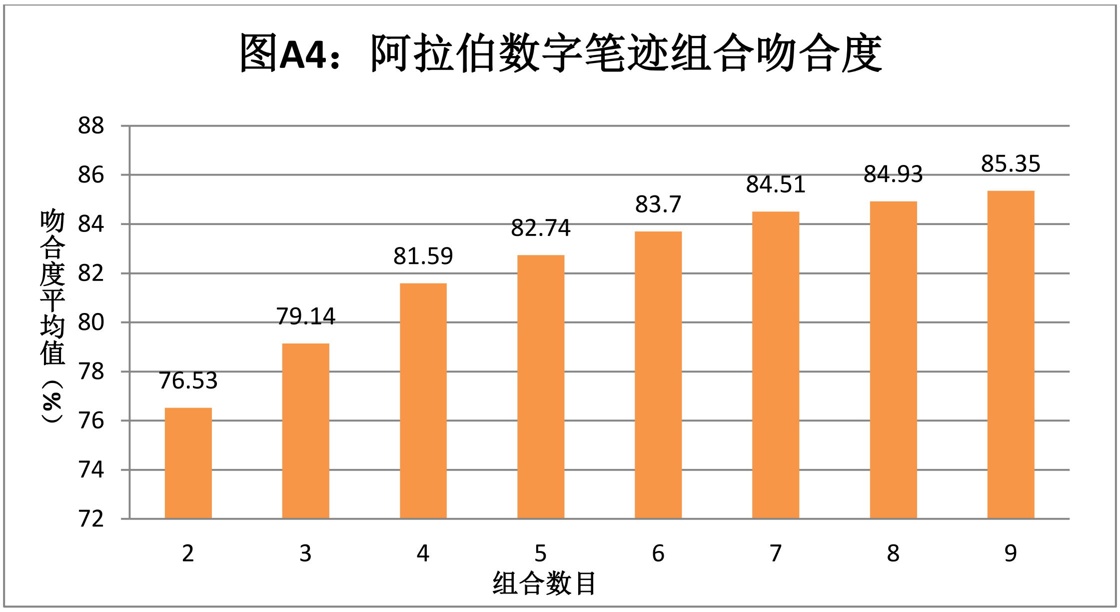
并观察随着组合数目递增，前后两次吻合度之间相对误差的变化。

此时，吻合度的定义仍为

$$吻合度 = \frac{吻合的特征数}{比较的总特征数} \times 100\%$$

表 4：阿拉伯数字笔迹组合吻合度

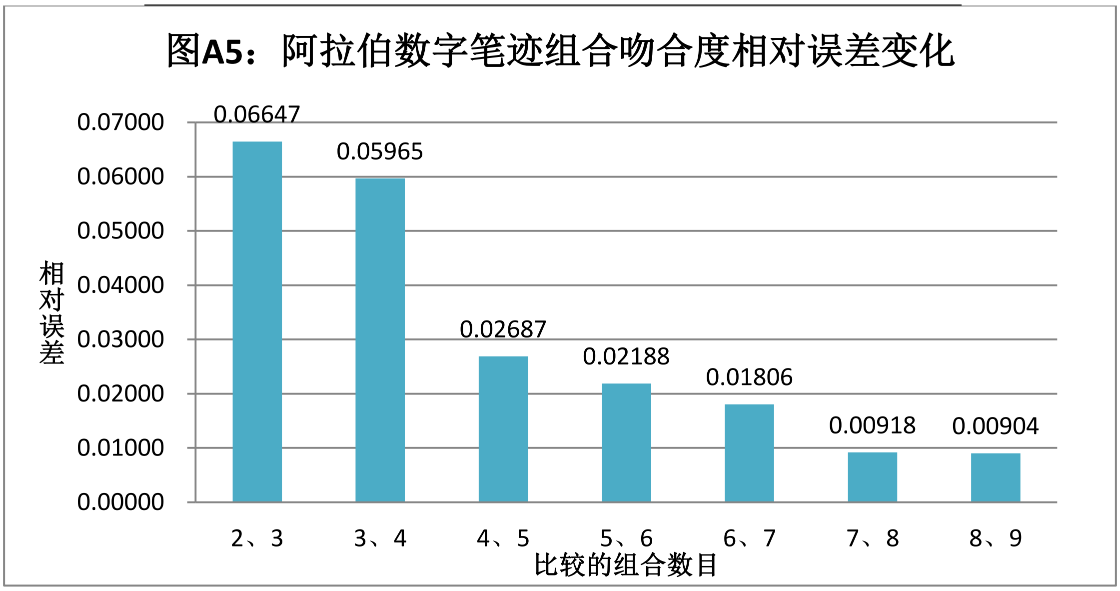
组合数目	平均吻合度（%）
2	76.53
3	79.14
4	81.59
5	82.74
6	83.7
7	84.51
8	84.93
9	85.35



由表 4 及其相对应的图 A4 可知，随着组合在一起阿拉伯数字数目的增多，吻合度增加并趋于稳定，最开始随着组合数目的增加，吻合度的增加也较快，而后增加的速度也趋于平缓。

表 5：阿拉伯数字笔迹组合吻合度相对误差变化

比较组合数目	相对误差
2、3	0.06647
3、4	0.05965
4、5	0.02687
5、6	0.02188
6、7	0.01806
7、8	0.00918
8、9	0.00904



由表 5 及其对应的图 A5 可知，随着组合数目的增加，前后两次吻合度之间的相对误差先是急剧下降，后趋于稳定，在组合数目为 7 时达到稳定（相对误差$\alpha < 0.02$）。

3.4 阿拉伯数字笔迹不同特征相对离散性比较

在所提取的特征中，主要分为三大类，角度（包括与水平面的倾斜角和夹角）、曲率、比例（包括上下部分比例和左右部分比例）。

用相对离散性来比较三种不同种类特征的稳定性。

将所有特征分类后，计算出三种不同种类特征的样本平均值，再以平均值为基础计算出标准差。

样本平均值定义：$\overline{X} = \frac{X_1+X_2+\cdots+X_n}{n}$

标准差定义：$s = \sqrt{\frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\cdots+(X_n-\overline{X})^2}{n}}$

随后以此为基础计算出变异系数(标准差与样本平均值的比率)：

变异系数：$CV = \frac{s}{\overline{X}}$ 以此能反映出三种不同种类特征在单位均值上的离散程度，从而比较判断出它们各自的稳定性。

表,6：阿拉伯数字特征样本 1 相对离散性分析

	1.1	1.2	1.3	平均值	标准差	变异系数
角度	31	28	28	29.00	1.73	0.02
	63	66	72	67.00	4.58	0.05
	75	78	76	76.33	1.53	0.01
	36	38	34	36.00	2.00	0.06
	13	16	14	14.33	1.53	0.07
	40	39	49	42.67	5.51	0.12
	56	50	53	53.00	3.00	0.03
	35	34	37	35.33	1.53	0.04
	79	78	78	78.33	0.58	0.00
	36	37	33	35.33	2.08	0.06
	62	60	56	59.33	3.06	0.04
	70	66	69	68.33	2.08	0.02
	78	75	74	75.67	2.08	0.01
	30	25	25	26.67	2.89	0.04
	73	80	75	76.00	3.61	0.03
	73	70	72	71.67	1.53	0.01
	75	53	39	55.67	18.15	0.16
曲率	3.84	3.78	4.04	3.89	0.14	0.03
	2.58	2.97	2.77	2.77	0.20	0.04
	2.89	2.3	2.7	2.63	0.30	0.08

	5.05	2.46	3.23	3.58	1.33	0.16
	2.24	2.48	2.29	2.34	0.13	0.04
	4.24	3.63	3.44	3.77	0.42	0.04
	3.05	5.25	3.43	3.91	1.18	0.24
	3.24	2.92	4.47	3.54	0.82	0.22
	3.28	2.96	2.59	2.94	2.00	0.68
	2.46	2.59	2.22	2.42	0.19	0.08
	1.95	2.85	2.34	2.38	0.45	0.12
	2.11	2.16	3.38	2.55	0.72	0.24
	2.44	1.56	2.6	2.20	0.56	0.24
	3.62	2.7	3.07	3.13	0.46	0.07
	2.82	3.2	3.42	3.15	0.30	0.05
	5.21	2.38	5.62	4.40	1.76	0.37
	3.5	2.13	4.11	3.25	1.01	0.31
	3.34	2	2.48	2.61	0.68	0.12
	3.68	3.21	2.78	3.22	0.45	0.08
	3.36	3.88	4.27	3.84	0.46	0.06
	3	3.72	3.31	3.34	0.36	0.07
	4.25	2.63	3.18	3.35	0.82	0.11
	4.39	3.78	3.18	3.78	0.61	0.09
	2.9	3.03	3	2.98	0.07	0.01
	2.35	2.09	3.73	2.72	0.88	0.30
	4.76	4	3.88	4.21	0.48	0.04
比例	0.95	0.72	0.78	0.82	0.12	0.06
	1.24	1.37	1.13	1.25	0.12	0.10
	1.8	1.52	0.71	1.34	0.57	0.32
	0.56	0.65	0.69	0.63	0.07	0.05
	1.19	0.76	1.38	1.11	0.32	0.28

图 B0：阿拉伯数字特征样本 1 相对离散性分析

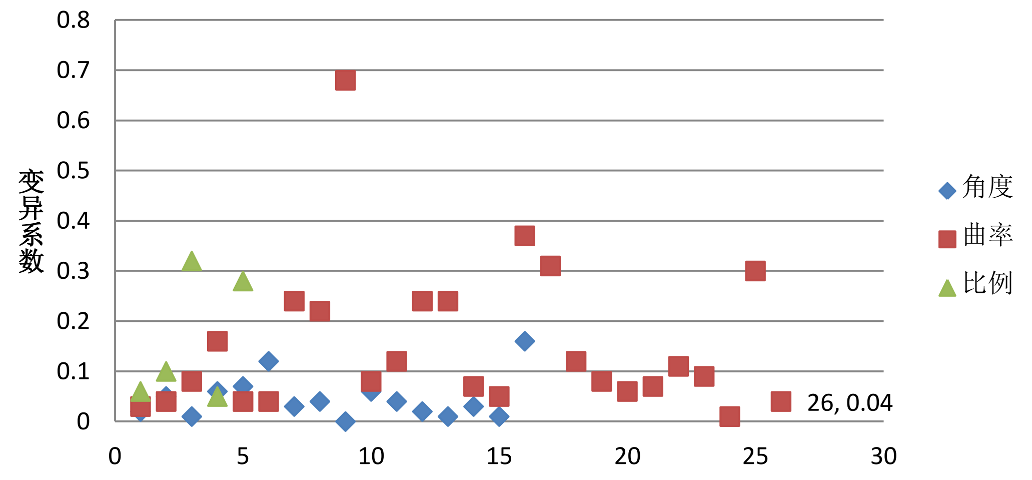
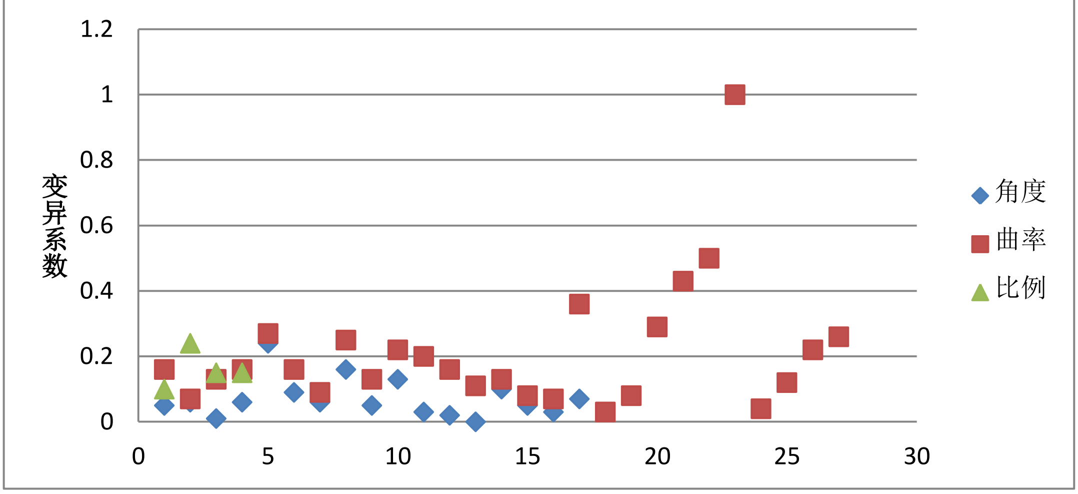


表 7：阿拉伯数字特征样本 2 相对离散性分析

	2.1	2.2	2.3	平均值	标准差	变异系数
角度	43	41	45	43.00	2.00	0.05
	67	73	65	68.33	4.16	0.06
	88	90	88	88.67	1.15	0.01
	45	40	41	42.00	2.65	0.06
	11	7	8	8.67	2.08	0.24
	45	50	54	49.67	4.51	0.09
	47	42	47	45.33	2.89	0.06
	43	42	32	39.00	6.08	0.16
	71	78	77	75.33	3.79	0.05
	38	43	33	38.00	5.00	0.13
	69	66	66	67.00	1.73	0.03
	80	77	78	78.33	1.53	0.02
	71	71	71	71.00	0.00	0
	28	23	25	25.33	2.52	0.1
	73	76	81	76.67	4.04	0.05
	78	75	74	75.67	2.08	0.03
	46	41	41	42.67	2.89	0.07
曲率	2.33	3.08	3.2	2.87	0.47	0.16
	3.63	3.55	3.17	3.45	0.25	0.07
	3.46	2.8	2.75	3.00	0.40	0.13
	5.38	3.88	5	4.75	0.78	0.16
	3	2.57	4.29	3.29	0.90	0.27
	3.41	3.08	4.21	3.57	0.58	0.16
	2.3	2.71	2.69	2.57	0.23	0.09
	2.29	3.45	3.84	3.19	0.81	0.25
	4.5	3.45	4	3.98	0.53	0.13
	3.29	2.21	3.42	2.97	0.66	0.22
	1.96	2.95	2.5	2.47	0.50	0.2
	0.81	1	1.11	0.97	0.15	0.16
	2.46	2.56	3	2.67	0.29	0.11
	2.41	2.74	3.12	2.76	0.36	0.13
	3.73	3.33	3.18	3.41	0.28	0.08
	3.53	3.1	3.21	3.28	0.22	0.07
	2.5	3.54	5.13	3.72	1.32	0.36
	2.2	2.24	2.35	2.26	0.08	0.03
	4.8	4.1	4.35	4.42	0.35	0.08
	4.58	3.48	2.52	3.53	1.03	0.29
	2.43	0.96	2.35	1.91	0.83	0.43
	2.96	5.29	1.95	3.40	1.71	0.5
	2.19	2.24	1.68	2.04	2.04	1
	3.18	3.18	3.41	3.26	0.13	0.04
	3.52	3.16	4	3.56	0.42	0.12
	3.05	4.73	4.28	4.02	0.87	0.22
	2.26	2.71	3.75	2.91	0.76	0.26
比例	0.62	0.7	0.58	0.63	0.06	0.1
	1.59	0.97	1.38	1.31	0.32	0.24

	2.2	2.6	1.94	2.25	0.33	0.15
	0.53	0.71	0.62	0.62	0.09	0.15



图B1：阿拉伯数字特征样本2相对离散性分析

由表 6、表 7 及相应的图 B0、图 B1 可知，不同形式的特征的稳定性不同且有一定的差异。目前的特征种类有曲率（提取数字弯曲处的特征）倾斜角，夹角（提取线条较笔直光滑的转折处的特征）、各部分的大小比例。根据变异系数来看，角度的变异系数最小，在单位均值上的离散程度最小，稳定性最高；曲率和比例的变异系数次之，且二者相当，在单位均值上的离散程度比角度高，即稳定性比角度差。阿拉伯数字笔迹中，角度的稳定性较好，且区分度较大，曲率、上下，左右部分比例有一定的稳定性，次之于角度。

3.5 本鉴定方法有效性分析

3.5.1 有效性分析：增加样本比较特征数

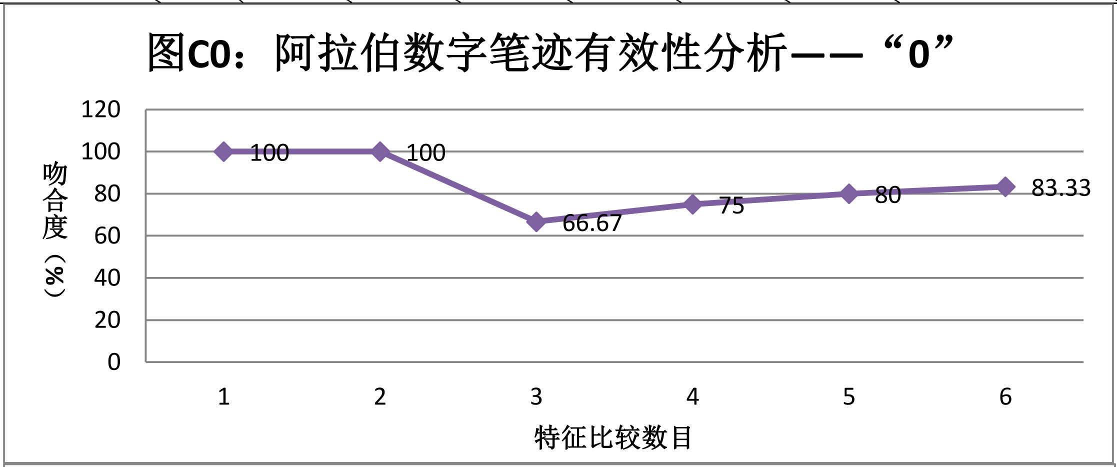
对同一个人书写的每个阿拉伯数字，不断增加进行比较的特征的数目，观察吻合度是否趋近于并稳定在一定范围，并得出使每个阿拉伯数字的吻合度开始趋于稳定的特征数。

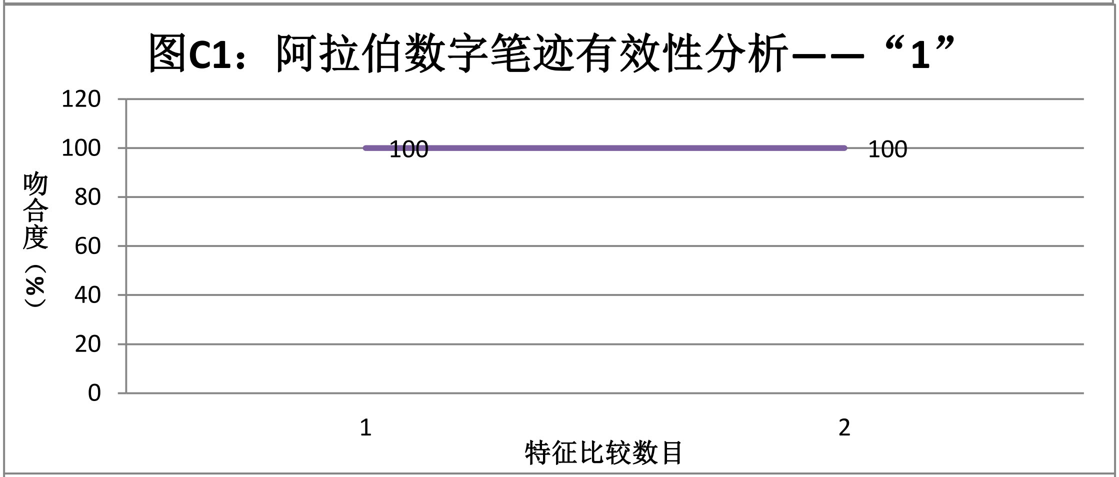
下表中，横行代表比较的阿拉伯数字，纵列代表进行比较的特征数目，列出的数据表示这种情况下的吻合度（%），标粗的部分为使每个阿拉伯数字的吻合度开始趋于稳定（定义为前后两次吻合度的相对误差α ＜ 0.1）的时候的吻合度。

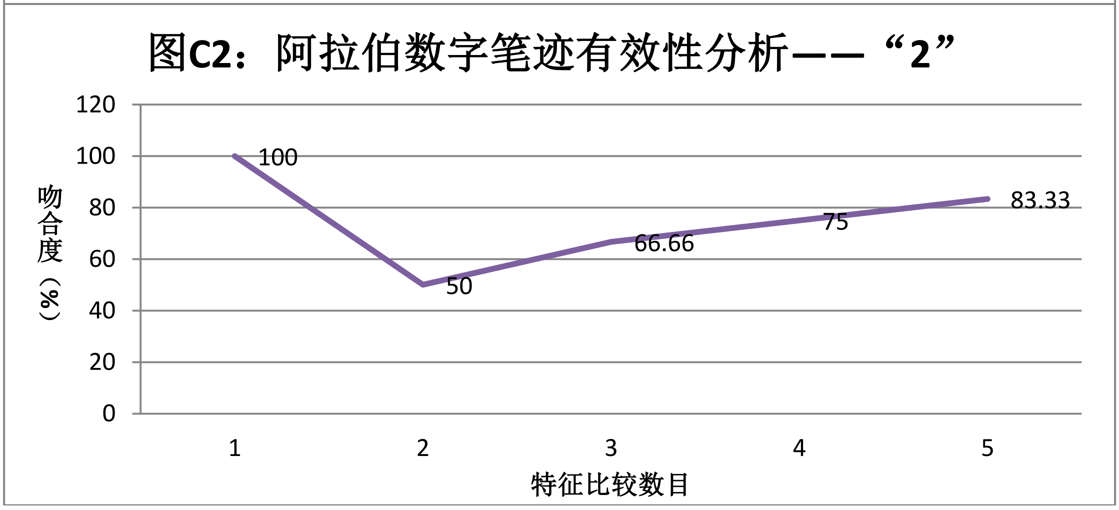
表 8：阿拉伯数字笔迹有效性分析：增加样本比较特征数

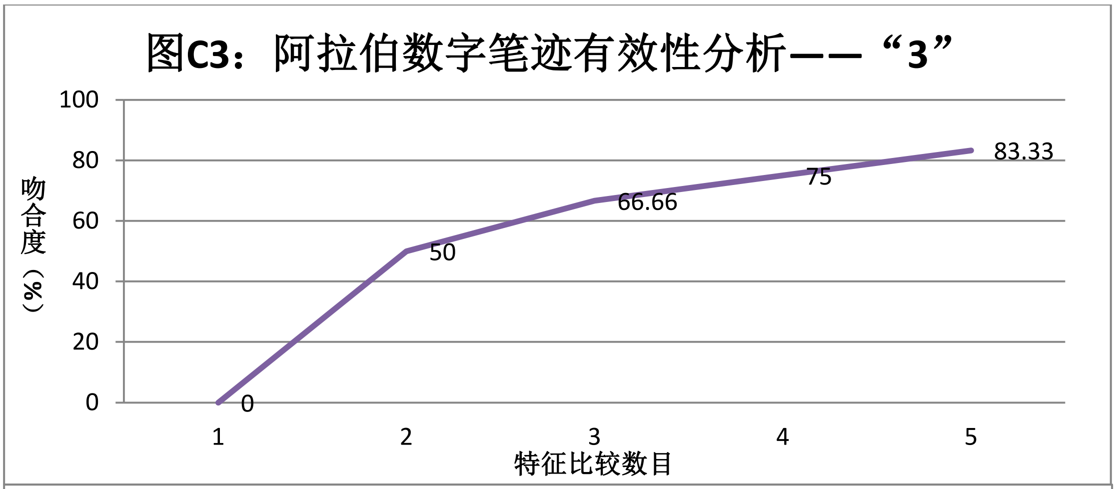
	0	1	2	3	4	5	6	7	8	9
1	100	**100**	100	0	100	100	100	100	100	100
2	100	100	50	50	50	100	100	50	50	100
3	66.67		66.66	66.66	**66.67**	66.67	66.67	**66.67**	66.67	66.67

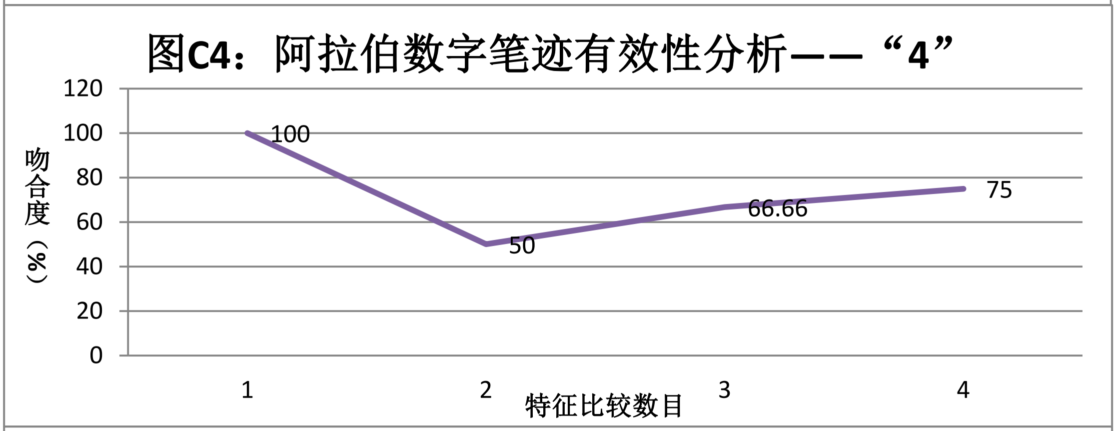
4	75		**75**	**75**	75	**75**	75		75	**75**
5	**80**		83.33	83.33		83.33	**80**		60	83.33
6	83.33						83.33		66.67	
7									**71.43**	
8									75	

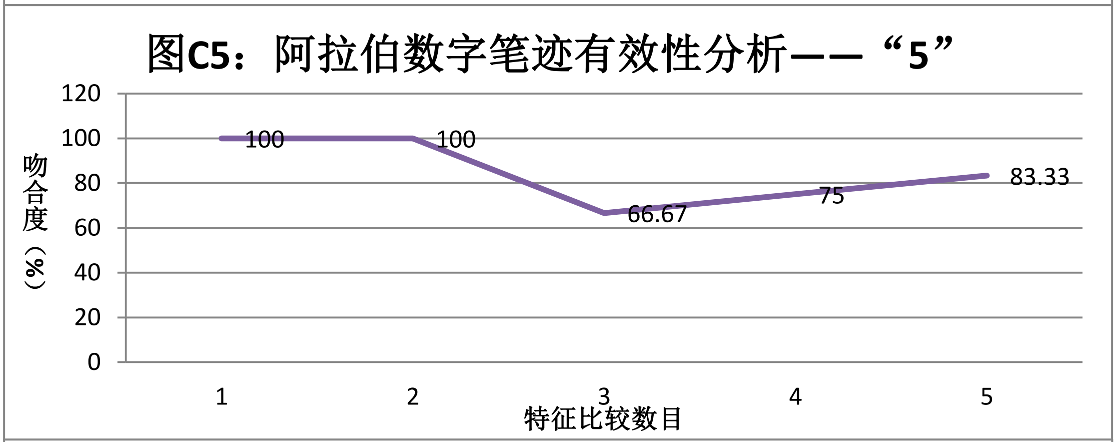


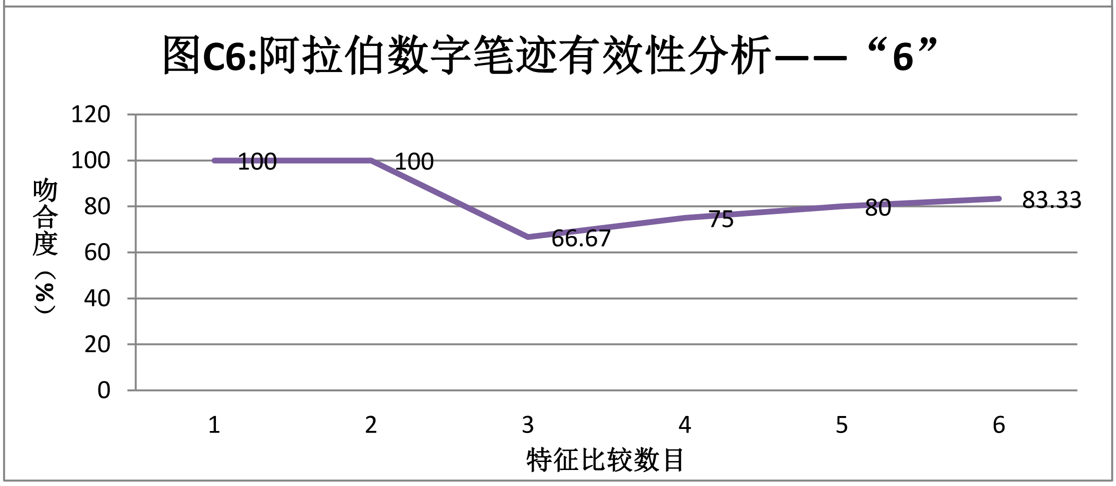


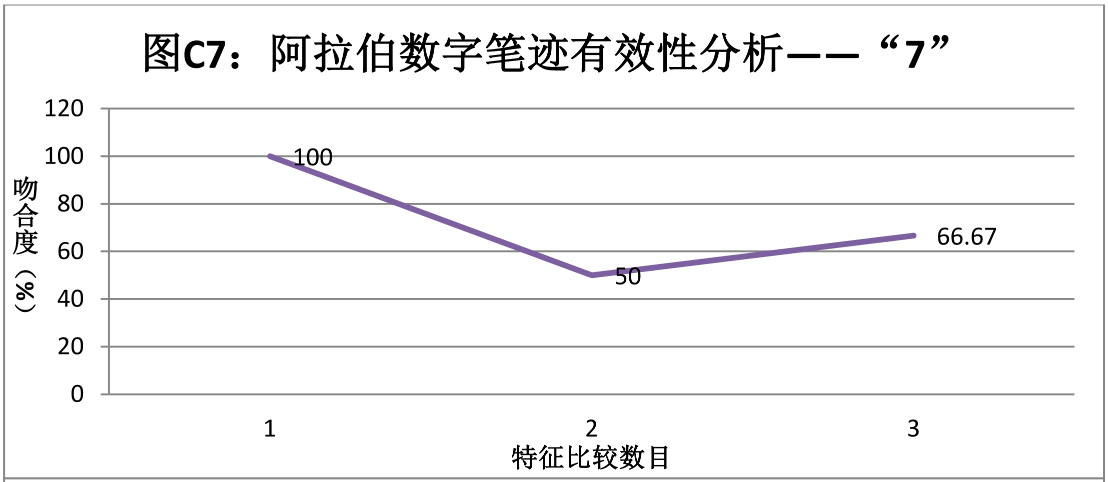


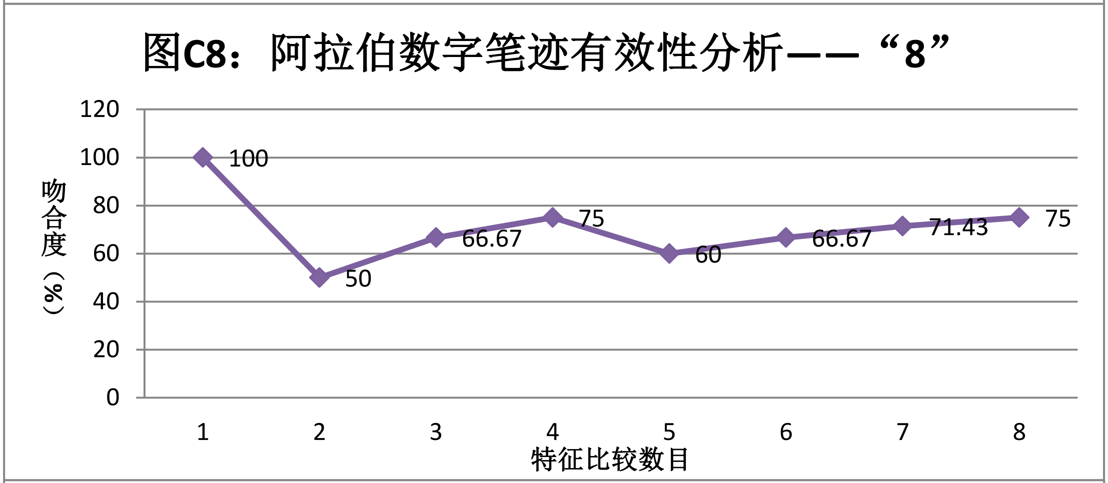

图C3：阿拉伯数字笔迹有效性分析——“3”
吻合度（%）
100
80
60
40
20
0
0
50
66.66
75
83.33
1
2
3
4
5
特征比较数目

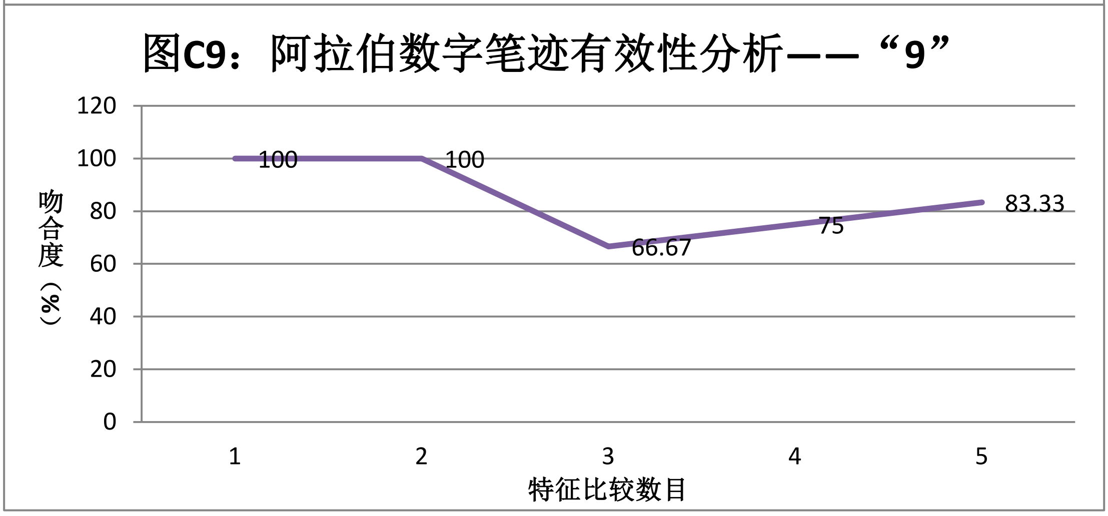

图C4：阿拉伯数字笔迹有效性分析——“4”
吻合度（%）
120
100
80
60
40
20
0
100
50
66.66
75
1
2
3
4
特征比较数目


图C5：阿拉伯数字笔迹有效性分析——“5”
吻合度（%）
120
100
80
60
40
20
0
100
100
66.67
75
83.33
1
2
3
4
5
特征比较数目


图C6:阿拉伯数字笔迹有效性分析——“6”
吻合度（%）
120
100
80
60
40
20
0
100
100
66.67
75
80
83.33
1
2
3
4
5
6
特征比较数目







由表 8 和上图对同一个人书写的每个阿拉伯数字，不断增加进行比较的特征的数目的有效性分析可知，随着进行比较的特征数目的增加，吻合度先是有所波动，最终都将趋近于并稳定在一定范围内，证明误差也随之减小并稳定在一定范围内，且使每个阿拉伯数字的吻合度开始趋于稳定的特征数量是不同的，使 0 至 9 这十个阿拉伯数字的吻合度开始趋于稳定的特征数量分别是 5、1、4、4、3、4、5、3、7、4。

3.5.2 有效性分析：增加提取样本特征并进行比较的次数

选取特征较多、具有代表性的阿拉伯数字，不断增加提取样本特征并进行比较的次数，观察吻合度是否趋近于并稳定在一定范围内。本课题中选取了阿拉伯数字 0、6 作为进行这方面有效性分析的数字，其中 0 有六个特征，6 有六个特征。

表 9：阿拉伯数字笔迹有效性分析——“6”

比较次数	1	2	3	4	5	6
吻合度（%）	66.67	83.33	50	66.67	83.33	83.33

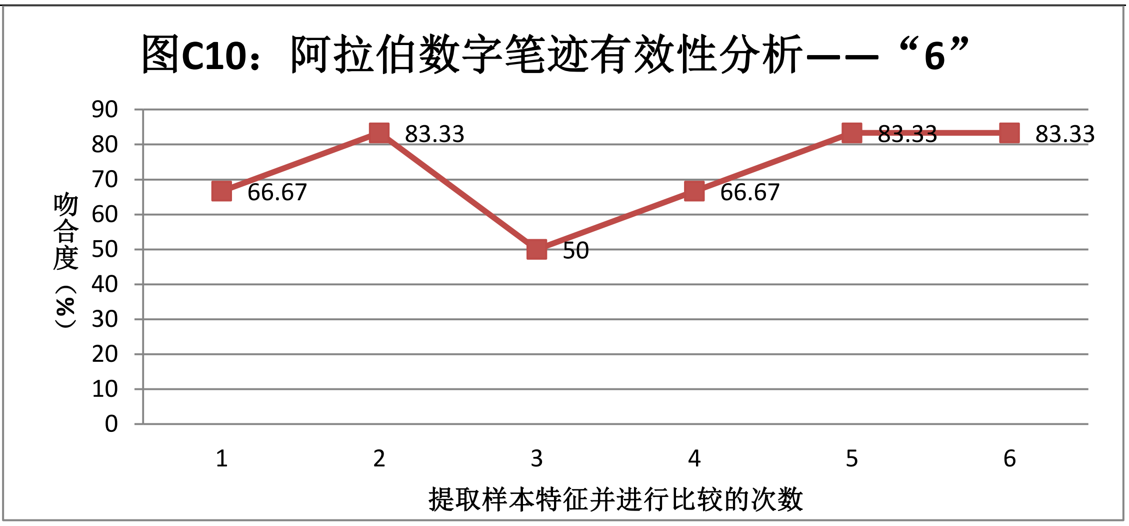
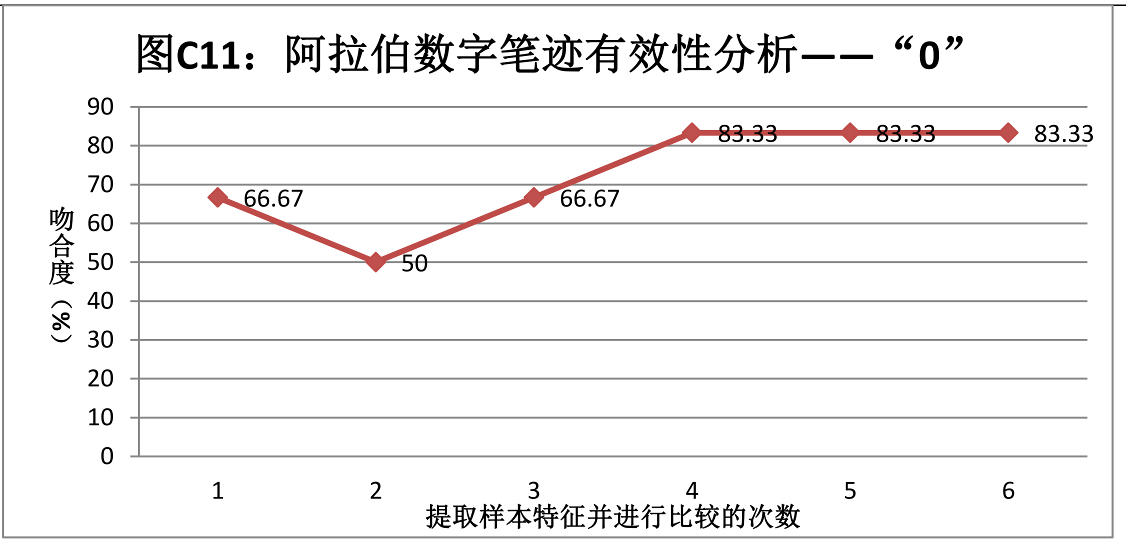


表 10：阿拉伯数字笔迹有效性分析——“0”

比较次数	1	2	3	4	5	6
吻合度（%）	66.67	50	66.67	83.33	83.33	83.33



由表 9、表 10 及其相对应的图 C10、图 C11 对阿拉伯数字 0 和 6 的有效性分析可知，随着提取样本特征并进行比较的次数的增多，吻合度先是有所波动，最终都将趋近于并稳定在一定范围内，证明误差也随之减小并稳定在一定范围内，本课题中使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法具有有效性、可重复性，在日后真正的鉴定领域内可以作为应用的参考。

3.6 测试

选取阿拉伯数字“8”作为测试的对象，共有八个特征。

分别令两人书写阿拉伯数字“8”三遍，分别提取阿拉伯数字“8”的八个特征，提取并计算出关于该数字特征的平均值作为其关于该数字的特征的数据，计算各个特征的标准差，进一步验证选取的特征是否稳定具有科学性以便进行调整。

表 11：阿拉伯数字笔迹研究方法测试——样本 1

样本 1.1	样本 1.2	样本 1.3	平均值	标准差
3.34	2	2.48	2.61	0.68
3.68	3.21	2.78	3.22	0.45
1.19	0.76	1.38	1.11	0.32
3.36	3.88	4.27	3.84	0.46
3	3.72	3.31	3.34	0.36
4.25	2.63	3.18	3.35	0.82
4.39	3.78	3.18	3.78	0.61
73	80	75	76.00	3.61

表 12：阿拉伯数字笔迹研究方法测试——样本 2

样本 2.1	样本 2.2	样本 2.3	平均值	标准差
4.8	4.1	4.35	4.42	0.35
4.58	3.48	2.52	3.53	1.03
0.53	0.71	0.62	0.62	0.09
2.43	0.96	2.35	1.91	0.83
2.96	5.29	1.95	3.40	1.71
2.19	2.24	1.68	2.04	0.31
3.18	3.18	3.41	3.26	1.03
73	76	81	76.67	2.04

由表 11、表 12 得出样本 1、2 三次样本的平均值作为其关于阿拉伯数字“8”的特征与之后所书写的进行比较。由计算所得的标准差可知，大部分特征的标准差稳定在 1 以内，有 2~3 个特征的标准差出现了异常，这几个特征之后需要进行进一步的调整。

再令两个人分别又书写六遍该数字，用比较吻合度的方法判断这些数字分别是谁书写的。目前，十二次判断中判断对了十次，正确率约为 83.33%。

可初步验证本课题中使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法具有有效性，在日后真正的鉴定领域内可以作为应用的参考。

3 结论与讨论

3.1 结论

（1）同一个人书写的样本中，吻合度大致稳定在 80%—100%之间，不同人书写的样本中，吻合度大致稳定在 20%—40%之间，二者间村子显著差异，可以利用特征的提取、吻合度的计算将其区分开来。随着阿拉伯数字组合数目的增加，吻合度变化并趋于稳定，最开始随着吻合度的增加变动较快，而后变化的速率也趋于平缓。

（2）不同形式的特征的稳定性不同。目前的特征种类有曲率（提取数字弯曲处的特征）倾斜角，夹角（提取线条较笔直光滑的转折处的特征）、各部分的大小比例。根据变异系数来看，角度的变异系数最小，在单位均值上的离散程度最小，稳定性最高；曲率和比例的变异系数次之，且二者相当，在单位均值上的离散程度比角度高，稳定性比角度差。一个阿拉伯数字笔迹中，角度的稳定性较好，且区分度较大，曲率、上下，左右部分比例有一定的稳定性，次之于角度。

（3）经过有效性分析，随着进行比较的特征数目的增加，吻合度都将趋近于并稳定在一定范围内，且使每个阿拉伯数字的吻合度开始趋于稳定的特征数量是不同的，使 0 至 9 这十个阿拉伯数字的吻合度开始趋于稳定的特征数量分别是 5、1、4、4、3、4、5、3、7、4；随着提取样本特征并进行比较的次数的增多，吻合度都将趋近于并稳定在一定范围内。都说明在这种鉴定方法中，误差终将减小并稳定在一定范围内，本课题中使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法具有有效性、可重复性，在日后真正的鉴定领域内可以作为参考。

3.2 课题讨论

阿拉伯数字笔迹和广为人们所熟知的汉字不同，它的主要特征包括笔画少、字体具有多样性、比较容易模仿等。大多数阿拉伯数字字形简单、笔画较少，这也导致了相比汉字而言，其更加容易被犯罪分子所模仿。但凡阿拉伯数字出现的案件中，一般都牵连到大额金额的案件，比如财产纠纷等，其中的经济方面的影响十分巨大和广泛，不得不引起人们的关注。

阿拉伯数字虽然简单，且形态方面的特征可能会一定程度上受到书写环境等客观因素的影响，根据之前系统的文献研究，阿拉伯数字笔迹的书写具有一定的动力定型，书写的动作习惯在一定时间内保持相对稳定，是阿拉伯数字笔迹统一认定的前提。而且，也正因为阿拉伯数字笔迹的简单，容易受到犯罪分子的忽视，也就给了案件一定的突破口。

笔迹检验鉴定，顾名思义，就是利用笔迹特征进行检验，得出关于笔迹的正确结论。阿拉伯数字笔迹的检验鉴定，有其自己的独特性，也有和汉字、外国字符笔迹鉴定方面相通的共性。

本课题以阿拉伯数字笔迹作为研究对象，给出的主要思路和方法是，提取阿拉伯数字笔迹数学方面的特征，将每个阿拉伯数字量化成一定量的数据。阿拉伯数字笔迹比较明显的特征：数字有弯曲的部分（例如 0、2、3、5、6、8、9 均成立）提取曲率作为特征、线条较笔直光滑的转折处（例如 2、4、5、7 均成立）提取倾斜的角度为特征、也包括各部分的大小比例等。建立好阿拉伯数字笔迹特征库后，运用统计分析的方法，计算出吻合度来判断是否为同一人书写。经过一系列的有效性分析和相关的检验证明，本课题中使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法具有有效性、可重复性，在日后真正的鉴定领域内可以作为参考。

在面对阿拉伯数字笔迹模仿问题中，首先提出阿拉伯数字笔迹的数学具有动力定型，尽管刻意模仿，但各个方面的特征仍会稳定在一定范围内，且根据文中不同种类特征的相对离散性分析得出角度是其中最稳定的，这个时候可以多提取一些角度作为比较的特征。而且，在可以模仿的状态下，书写的速度、力度均会受到影响，也可以作为研究的一个指标。当然，这种情况下，最重要的还是要和鉴定人的经验相结合，以期得出最准确的结论。

运用数学统计方面的方法来处理阿拉伯数字笔迹的鉴定，可以去除主观经验性主观方面较大的影响，用一种科学具有有效性、可重复性的方法来研究这个问题。且在最后综合分析的阶段，可以结合一些自己的经验方面的因素来更好地分析这个问题，做出最准确的判断。

在面对具体案件的时候，也应该具体分析，本着科学严谨的态度，准确收集样本、分析阿拉伯数字笔迹的特征、结合案情的实际情况来具体分析，才能得到最为准确的结论。本课题中的研究方法也仅仅是给出了一种实践的思路，希望综合日后可能的实践，进行进一步的发展和改进，以期得到真正可行有效的应用。

3.3 创新点

（1）首次跨过表面的形态方面的观察和经验层面，用特征提取、计算吻合度的定量的判断方法，对阿拉伯数字笔迹的特征提取分析及其鉴定应用进行了研究，实现了从定性到定量的过程。为今后阿拉伯数字笔迹鉴定实践过程中提供一定的理论依据支撑和可供参考的新思路，这一点是本课题最大的实际应用。除此之外，这项研究项目也能让大家对这个平时不是很熟知的领域有一些新的了解。

（2）进行了有效性分析与测试，都说明在这种鉴定方法中，误差终将减小并稳定在一定范围内，本课题中使用的对阿拉伯数字笔迹的特征提取及吻合度的分析方法具有有效性、可重复性，在日后真正的鉴定领域内可以作为参考。

3.4 课题展望

（1）应用计算机编程实现自动化提取特征进行比较、计算吻合度的过程，目前只是应用了计算机测量软件，提取特征速度较慢、耗时，希望在速度方面有所提高，达到快速、准确的信息化过程。

（2）增加样本的数目，进一步观察并验证用特征提取、计算吻合度的定量的判断方法，对阿拉伯数字笔迹的特征提取分析的有效性。

（3）和实际案例相结合，对本课题中使用的对阿拉伯数字笔迹的特征定量提取分析的方法在鉴定领域内的实际应用进行探讨。

（4）从阿拉伯数字笔迹更复杂的特征方面，包括阿拉伯数字的连笔、在原有阿拉伯数字上的涂改，添减局部特征、阿拉伯数字落笔的力度、速度等方面来进行进一步的更深入的分析。

致谢

感谢母校给我提供课题研究的场所、条件；

感谢专家和助教对我的课题提供的指导和改进意见；

感谢我的指导老师在课题研究中对我进行了全程的指引；

感谢我的父母给了我许多鼓励和支持；

感谢耐心认真帮助我完成样本收集的受访者！

参考文献

[1]徐晗，阿拉伯数字笔迹鉴定研究 西南政法大学.2011 年 3 月

[2]道格拉斯·A·林德、威廉·G·马歇尔、罗伯特·D·梅森，商务与经济统计技术（第11版） 中国人民大学出版社.2005年9月

[3]吴琦，浅析阿拉伯数字笔迹的检验方法 《森警科技》.2006年第6期

[4]王连昭，阿拉伯数字笔迹检验探析 《犯罪研究》.2012年第1期

[5]王连昭，阿拉伯数字笔迹特征新探 《福建警察学院学报》.2012年第4期

[6]董慧格，笔迹样本的收集方法 《科技风》.2010第24期

[7]黄祥，浅析摹仿签名笔迹鉴定 《中国司法鉴定》.2004年第3期

[8]李建明，一起单个阿拉伯数字添改笔迹鉴定的体会 广东省梅州市公安局文检室

[9]王连昭，阿拉伯数字笔迹鉴定科学基础的理论探索 《吉林公安高等专科学校学报》.2012年第4期

[10]程连栋，“转折角度”检验阿拉伯数字 《刑事技术》.1996年第1期

附：收获与体会

这是自己第一次接触数学方面的课题，能完成下来感觉也是十分不容易。

本来只是一个偶然生出的想法，查阅了大量文献发现这一块确实有很多东西可以做的时候也是十分兴奋。但没想到第一步便遇到了难题，便是如何提取大家所书写的阿拉伯数字笔迹的特征。记得当时上网搜寻了大量测量软件，但大多都无法满足自己需要测量曲率的需求，最终功夫不负有心人，在多方询问之后决定用 Image-J 这个软件来完成测量，不仅简单易学，效果也是十分不错的，但是提取特征还是比较慢、耗时，所以希望以后能进一步改善，达到自动化的过程。

确定到底提取阿拉伯数字笔迹的哪些特征那块也是费了不少的脑筋，找了许多数学老师讨论，也在同学、家长不同人群中收集了大量初步的样本，仔细观察看上去差异较大之处，最终才算初步敲定准备提取的特征。第一次是设计了 26 个特征，到后来真的作比对分析发现太少时又增加到了 48 个特征。

关于特征提取后的分析方面，本来是 0 至 9 十个阿拉伯数字综合在一起算吻合度进行不同的比对的，后来经过老师的指导发现应该把 0 至 9 十个阿拉伯数字分开来分别统计吻合度才算比较科学，毕竟真实案例中不可能 0 至 9 十个阿拉伯数字都出现，而且有可能一个数字的数据影响整体 0 至 9 十个阿拉伯数字的数据的结果。后来再想到，可以不断增加组合的数目，来观察和只有一个阿拉伯数字时吻合度的比较，两者结合结论总归更加丰满。

而后半部分的有效性分析以及最后的测试部分，本来并没有想到这部分,但做完阿拉伯数字笔迹特征提取后的统计分析时，被指导老师启发应该要证明这种研究方法具有可重复性、科学性，研究方法有误差是必然的，但必须要找到一个角度分析判断误差是否迅速趋近于零，否则这个方法便是无效、不可重复的。受到这个启发后，我便从不同角度设计了有效性分析以及最后的测试，最终也证明这种研究方法是科学可行有效的。

经过这个数学课题，不仅强化了自己在数学理论知识基础上对实际问题的分析、应用能力，更让自己对数学产生了兴趣。目前，自己已经决定在大学主修统计专业，也已经在专门自学相关更加系统的知识，想必让自己发现兴趣、实践兴趣也是做课题作重要的意义吧。